JAMA Surgery | Original Investigation

Development and Validation of the Scoring System of Appendicitis Severity 2.0

Jochem C. G. Scheijmans, MD; Wouter J. Bom, MD, PhD; Umme Habiba Ghori, Bsc; Anna A. W. van Geloven, MD, PhD; Gerjon Hannink, PhD; Charles C. van Rossem, MD, PhD; Lieke van de Wouw, MD; Peter M. Huisman, MD; Annemiek van Hemert, MD; Rutger J. Franken, MD; Steven J. Oosterling, MD, PhD; Camiel Rosman, MD, PhD; Lianne Koens, MD, PhD; Jaap Stoker, MD, PhD; Marcel G. W. Dijkgraaf, PhD; Marja A. Boermeester, MD, PhD; for the SAS Collaborative Group

IMPORTANCE When considering nonoperative treatment in a patient with acute appendicitis, it is crucial to accurately rule out complicated appendicitis. The Atema score, also referred to as the Scoring System of Appendicitis Severity (SAS), has been designed to differentiate between uncomplicated and complicated appendicitis but has not been prospectively externally validated.

OBJECTIVE To externally validate the SAS and, in case of failure, to develop an improved SAS (2.0) for estimating the probability of complicated appendicitis.

DESIGN, SETTING, AND PARTICIPANTS This prospective study included adult patients who underwent operations for suspected acute appendicitis at 11 hospitals in the Netherlands between January 2020 and August 2021.

MAIN OUTCOMES AND MEASURES Appendicitis severity was predicted according to the SAS in 795 patients and its sensitivity and negative predictive value (NPV) for complicated appendicitis were calculated. Since the predefined targets of 95% for both were not met, the SAS 2.0 was developed using the same cohort. This clinical prediction model was developed with multivariable regression using clinical, biochemical, and imaging findings. The SAS 2.0 was externally validated in a temporal validation cohort consisting of 565 patients.

RESULTS In total, 1360 patients were included, 463 of whom (34.5%) had complicated appendicitis. Validation of the SAS resulted in a sensitivity of 83.6% (95% CI, 78.8-87.6) and an NPV of 85.0% (95% CI, 80.6-88.8), meaning that the predefined targets were not achieved. Therefore, the SAS 2.0 was developed, internally validated (C statistic, 0.87; 95% CI, 0.84-0.89), and subsequently externally validated (C statistic, 0.86; 95% CI, 0.82-0.89). The SAS 2.0 was designed to calculate a patient's individual probability of having complicated appendicitis along with a 95% CI.

CONCLUSIONS AND RELEVANCE In this study, external validation of the SAS fell short in accurately distinguishing complicated from uncomplicated appendicitis. The newly developed and externally validated SAS 2.0 was able to assess an individual patient's probability of having complicated appendicitis with high accuracy in patients with acute appendicitis. Use of this patient-specific risk assessment tool can be helpful when considering and discussing nonoperative treatment of acute appendicitis with patients.

Invited Commentary

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

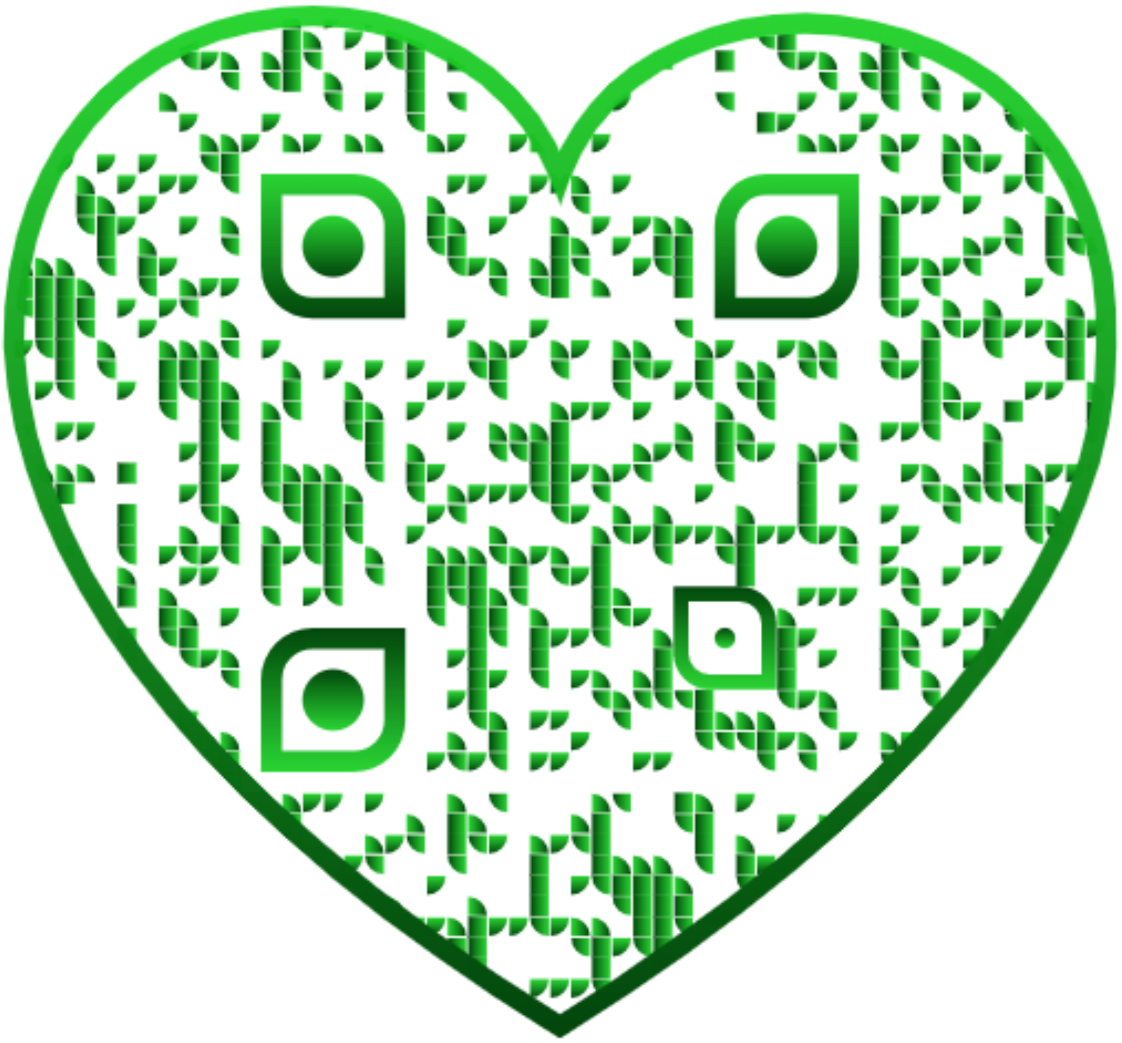
Group Information: The members of the SAS Collaborative Group appear in Supplement 2.

Corresponding Author: Jochem C. G. Scheijmans, MD, Department of Surgery, J1A, Amsterdam University Medical Center, Meibergdreef 9, Amsterdam 1105 AZ, the Netherlands (j.c.scheijmans@amsterdamumc.nl).

JAMA Surg. doi:10.1001/jamasurg.2024.0235
Published online March 27, 2024.

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Copenhagen University Library, Gabriel Wentz on 03/28/2024



Antibiotic treatment is a safe alternative to surgery for uncomplicated appendicitis.[1-4] To put nonsurgical treatment of acute uncomplicated appendicitis into practice, correct differentiation between uncomplicated and complicated appendicitis is essential.[5] However, this differentiation is difficult.[5,6] A meta-analysis of randomized clinical trials that attempted to include only patients with uncomplicated appendicitis based on clinical and imaging findings showed a complicated appendicitis rate of 16.9%.[7] More reliable ruling out of complicated appendicitis may minimize recurrent appendicitis after antibiotic treatment and therefore potentially lead to better treatment outcomes.[5]

Several scoring models have been developed to differentiate between uncomplicated and complicated appendicitis, incorporating both clinical and computed tomography (CT) findings.[8-14] Avanesov et al[13] developed a model with a positive predictive value (PPV) of 92% for complicated appendicitis, but its ability to accurately rule out complicated appendicitis—which is essential for treatment selection—is limited (negative predictive value [NPV] = 83%). On the other hand, Atema et al[14] composed a model for patients diagnosed by CT or ultrasonography (US), resulting in an NPV of 94.7% for CT and 97.1% for US. This score has been referred to as Scoring System of Appendicitis Severity (SAS),[5,15] and several studies have demonstrated the potential utility of this tool if externally validated.[11,16-18] Only 1 study has focused on the probability of complicated appendicitis, and this study was based on retrospective data and has not been widely referred to.[12]

The present study aims to externally validate the SAS in distinguishing uncomplicated from complicated appendicitis. Since target sensitivity and NPV of 95% each were not achieved, a new SAS (2.0) was developed and validated to estimate a patient's individual probability of having complicated appendicitis.

Key Points

Question How can complicated appendicitis be accurately ruled out when considering nonoperative treatment in patients with acute appendicitis?

Findings This multicenter study of 1360 patients with imaging-confirmed appendicitis found that the previously designed Scoring System of Appendicitis Severity (SAS) fell short in accurately ruling out complicated appendicitis. The newly developed SAS 2.0 was able to assess an individual patient's probability of having complicated appendicitis with high accuracy among patients with acute appendicitis.

Meaning In this study, the SAS 2.0 provided a safe basis for when to consider and discuss nonoperative treatment of uncomplicated appendicitis with patients.

Methods

An observational, prospective cohort study was conducted in 11 nonacademic referral hospitals in the Netherlands. The study design underwent formal approval by the institutional review board of Amsterdam University Medical Center, Amsterdam, the Netherlands, and was determined to be exempt from the regulations outlined in the Medical Research Involving Human Subjects Act owing to its observational nature. All participants provided written informed consent. This study was performed and reported according to the Standards for Reporting of Diagnostic Accuracy (STARD)[19] and TRIPODTransparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD)[20] reporting guidelines. The full study protocol has been published previously.[15]

Participants and Data Collection

All consecutive adult patients 18 years and older with a diagnosis of acute appendicitis based on clinical and imaging findings were approached for this study. Only patients who underwent surgery with the intention of having an appendectomy were included. Data on clinical, laboratory, and imaging findings were collected prospectively through standard reports in the electronic health record by attending physician in the emergency department, radiologist, surgeon, or pathologist, according to the study protocol.[15] All patients underwent imaging, and in cases where multiple modalities were involved (ie, magnetic resonance imaging or CT after US), data were collected from the decisive imaging modality, typically the last one performed.

SAS Validation

The primary outcome was the sensitivity and NPV of the SAS. Sensitivity was chosen because this study aimed to investigate the potential of the SAS for not missing cases of complicated appendicitis. NPV was chosen to investigate the potential of the SAS for ruling out complicated appendicitis. The original SAS consists of 2 separate scores, a US score and a CT score, depending on the last-performed imaging modality, as shown in eTable 2 in Supplement 1. An individual score was calculated for each patient. The score results in a preoperative diagnosis of uncomplicated appendicitis at 5 points or less for US and 6 or less for CT. In all patients, the anticipated diagnosis based on either the US or CT score was compared to the final diagnosis in order to validate the SAS. In this external validation, the predefined targets for sensitivity and NPV were both set at 95%, based on the generalized accuracy of the original SAS.[15] Since these targets were not achieved, an adjusted clinical prediction model we termed the SAS 2.0 was developed. Subsequently, the SAS 2.0 was externally validated in a second study cohort.

Reference Standard

The reference standard was the final diagnosis, based on surgical and pathological findings. Uncomplicated appendicitis was defined as inflammation or ulceration of the appendix.[21] Complicated appendicitis was defined as appendiceal inflammation with signs of gangrene or perforation, or a large intraperitoneal abscess or infiltrate.[21] In case of discrepancy between intraoperative and histopathological findings, surgical findings were decisive, except for normal appendices found by the pathologist; here, practice deviated from the pretrial protocol

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Copenhagen University Library, Gabriel Wentz on 03/28/2024

based on improved insights. The final diagnosis was assigned by an adjudication committee based on all available data.[15]

When calculating diagnostic accuracy, a normal appendix was placed under the heading of uncomplicated appendicitis since the purpose of this appendicitis severity score is not to diagnose appendicitis but to estimate the severity of the disease. In line with this reasoning, urgent diagnoses other than appendicitis that required surgery, including appendiceal malignancy, were placed under complicated appendicitis in the analysis, as surgery is indicated for these situations just as for complicated appendicitis.

Sample Size

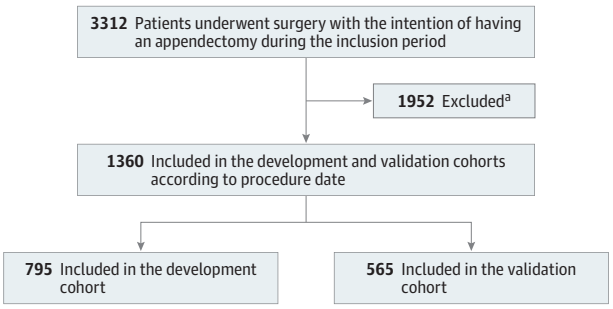
Patients were included in 2 cohorts. The first cohort comprised 795 patients and was used for validation of the original SAS as well as development of the SAS 2.0 (referred to as the development cohort). To ensure the target sensitivity and NPV of both to be 95% with a 3% lower margin set as the minimum of the 1-sided 97.5% adjusted Wald CI, a sample of 228 cases of complicated appendicitis were needed. Considering the prevalence of complicated cases to be 28.7% in the target population, the study protocol required 795 patients to ensure adequate validation of the original SAS.[15] This cohort size was also considered sufficient for the development of the SAS 2.0.[22] Thereafter, enrollment continued to form the validation cohort, which was used for external validation of the SAS 2.0. For this validation cohort, a lower limit of 5% rather than 3% was considered sufficient, resulting in a required number of at least 328 patients.

Statistical Analysis

For development of the SAS 2.0, both internal and external validation were performed. Odds ratios (ORs) for continuous variables were presented as IQR ORs with their 95% CIs, comparing a person with a typical high value (75th percentile) of the predictor to a person with a typical low value (25th percentile). All variables that were expected to predict complicated appendicitis based on the original SAS, other recent literature, or expert opinion were included in the analysis. Variables examined included age, sex, body temperature, numerical pain rating scale score, vomiting, duration of symptoms, white blood cell count, C-reactive protein (CRP) level, appendiceal diameter on imaging, and presence of any of the following on imaging: free intra-abdominal air or fluid, intra-abdominal abscess, appendicolith, fat infiltration, or appendiceal wall destruction. Missing data were analyzed and imputed using 25 imputation sets based on all other available other parameters.

For the development of the SAS 2.0, univariable and multivariable logistic regression models were used to investigate all included variables. In the final multivariable model, variables were selected based on Akaike information criterion.[20] Nonlinearity of continuous variables was tested with restricted cubic spline functions. Plausible or clinically meaningful interaction terms—that is, interactions between duration of symptoms and CRP, duration of symptoms and white blood cell count, duration of symptoms and temperature, and age and temperature—were tested and incorporated if significantly benefiting model fit. Internal validation of the prediction model was performed using bootstrap resampling with 500 replicates. Regression coefficients were pooled over the model fits for the multiple imputed datasets using the Rubin rule. For internal validation, bootstrap validations were performed separately for each model fit on each completed dataset, and the estimates of model performance were subsequently pooled. To improve the accuracy of the predictive model, the regression coefficients of the model were modified toward zero to reduce overfitting and improve generalizability, using the uniform shrinkage correction factor from the bootstrapping. Model performance was expressed in terms of discrimination and calibration.

Figure 1. Flow Diagram of Study Inclusions


3312 Patients underwent surgery with the intention of having an appendectomy during the inclusion period
1952 Excluded[a]
1360 Included in the development and validation cohorts according to procedure date
795 Included in the development cohort
565 Included in the validation cohort

[a] Did not meet inclusion criteria, met exclusion criteria, were not willing to participate, or logistic reasons. In most of these cases, no note was made in the file, so the exact reason for noninclusion is unknown. It was expected that the main reasons were logistic, such as forgetting to inform the patient about the study or to ask for participation.

Discrimination of the model was expressed by the concordance statistic (C statistic). Calibration was graphically presented with a flexible calibration plot to show whether the predicted diagnosis was in line with the observed diagnosis.[23] In addition, performance of the SAS 2.0 was investigated separately for subgroups of both patients diagnosed by US and CT. Finally, the SAS 2.0 was externally validated using the validation cohort. Since CRP determination is not standard in all hospitals worldwide, we incorporated the option to use the final model when the CRP value is unavailable.

For clinical applicability, an online decision SAS 2.0 tool was established. This tool can be used to calculate a patient's individual probability of having complicated appendicitis. Unlike the original SAS, the SAS 2.0 was developed without a predetermined cutoff value. However, the sensitivity and NPV of the SAS 2.0 tool were visually presented through a graph plotted against varying threshold levels of the probability of having complicated appendicitis. Statistical analyses were performed using SPSS version 26 (IBM) as well as R version 4.3.1 (R Foundation) and the rms package version 6.7-0.

Results

A total of 3312 adult patients underwent surgery for suspected acute appendicitis with the intention to perform an appendectomy, 1360 of whom were prospectively included

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Copenhagen University Library, Gabriel Wentz on 03/28/2024

Table 1. Baseline Characteristics by Cohort

Characteristic	No. (%)		
	Development cohort (n = 795)	Validation cohort (n = 565)	Original cohort (n = 395)[a]
Age, median (IQR), y	42 (29-57)	40 (27-56)	37 (27-50)
Sex			
Male	415 (52.2)	286 (50.6)	215 (54.4)
Female	380 (47.8)	279 (49.4)	180 (45.6)
Body temperature, mean (SD), °C	37.2 (0.8)	37.1 (0.8)	37.4 (0.8)
Duration of symptoms			
1 d (0-24 h)	372 (46.8)	277 (49.0)	NA[b]
2 d (24-48 h)	246 (30.9)	146 (25.8)	NA[b]
≥3 d (≥48 h)	177 (22.3)	142 (25.1)	NA[b]
White blood cell count, mean (SD), μL[c]	13 500 (4400)	13 400 (4600)	13 300 (10 500-16 400)[d]
C-reactive protein, median (IQR), mg/dL[e]	5.4 (2.1-10.1)	4.4 (2.0-8.6)	4.1 (1.6-9.5)
Last imaging			
Ultrasonography	472 (59.4)	367 (65.0)	111 (28.1)
Initial computed tomography	163 (20.5)	97 (17.2)	83 (21.0)
Computed tomography after inconclusive ultrasonography	160 (20.1)	101 (17.9)	NA
Ultrasonography and computed tomography[f]	NA	NA	201 (50.9)
Intraoperative diagnosis			
Uncomplicated appendicitis[g]	509 (64.0)	381 (67.4)	284 (71.1)
Complicated appendicitis[h]	286 (36.0)	184 (32.6)	114 (28.9)

Abbreviation: NA, not applicable.

[a] Original cohort refers to that in Atema et al.[14]

[b] Atema et al[14] only reported proportion of patients presenting with duration of symptoms ≥48 hours.

[c] To convert to $\times 10^9$/L, multiply by 0.001.

[d] Only medians with IQRs were included in Atema et al.[14]

[e] To convert to mg/L, multiply by 10.

[f] The cohort in Atema et al[14] consisted of 2 merged study cohorts, Optimization of Diagnostic Imaging Use in Patients With Acute Abdominal Pain (OPTIMA)[24] and Optimization of Imaging Appendicitis (OPTIMAP).[25] For the OPTIMA study, all patients underwent both ultrasonography and computed tomography regardless of the findings at ultrasonography or the clinically indication for direct computed tomography.

[g] Including normal appendicitis or other nonurgent diagnoses found during operation with intention to appendectomy.

[h] Including all other urgent diagnoses found during operation with intention to appendectomy (n = 7).

Table 2. Diagnostic Performance of the Scoring System of Appendicitis Severity (SAS) in External Validation[a]

Variable	No. (%)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Ultrasonography score	472 (59.4)	84 (76-90)	50 (44-55)	37 (32-43)	90 (84-94)
Computed tomography score	323 (40.6)	83 (77-89)	58 (50-66)	67 (60-73)	78 (69-85)
Total cohort	795 (100)	84 (79-88)	52 (48-57)	50 (45-54)	85 (81-89)

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

[a] External validation took place in a development cohort of 795 patients.

(**Figure 1**). The final diagnosis for these 1360 patients was a normal appendix in 15 patients (1.1%) and uncomplicated appendicitis in 875 (64.3%). Of the remaining 470 patients (34.6%) with an urgent diagnosis requiring surgery, 463 had complicated appendicitis, and 258 of these had perforated appendicitis (19%). Seven patients (0.5%) had an urgent diagnosis other than appendicitis, including intra-abdominal abscesses with peritonitis and cecal perforations. In the total cohort, 22 appendiceal neoplasms (1.6%) were found. The proportion of complicated appendicitis was 25.1% (211 of 839) in patients who were diagnosed by US compared to 49.7% (259 of 521) in patients diagnosed by CT.

Validation of the Original SAS

The original SAS was validated in a cohort of 795 patients (**Table 1**[24,25]). There were 472 patients diagnosed by US, 193 of whom had a score below the cutoff of 6 points, meaning a predicted diagnosis of uncomplicated appendicitis. Of these 193 patients labeled by the SAS as uncomplicated, 20 patients (10.4%) had complicated appendicitis according to the reference standard. Of 279 patients with a score of 6 points or higher, 104 had complicated appendicitis. Of the 323 patients with a CT scan, 121 had a score below the cutoff of 7 points, and 27 of these (22.3%) who were labeled by the SAS as uncomplicated had complicated appendicitis according to the reference stan-

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Copenhagen University Library, Gabriel Wentz on 03/28/2024

dard. Of 202 patients with a score of 7 points or higher, 135 had complicated appendicitis (66.8%). Diagnostic performance of both the US and CT scores is shown in **Table 2**. Combining both scores, the SAS had an overall sensitivity of 83.6% (95% CI, 78.8-87.6) and an NPV of 85.0% (95% CI, 80.6-88.8) for complicated appendicitis among patients with appendicitis. The specificity and PPV were 52.5% (95% CI, 48.0-56.9) and 49.7% (95% CI, 45.1-54.3), respectively. Thereby, the performance of the original SAS was below the predefined sensitivity and NPV threshold of at least 95%.

Development of the SAS 2.0

Univariable analysis of all included variables is shown in **Table 3**. Variable selection resulted in the following variables to be included in the SAS 2.0 model: sex, age, body temperature, numerical pain rating scale score, CRP level, appendiceal diameter, and radiological presence of free intra-abdominal fluid, an intra-abdominal abscess, an appendicolith, fat infiltration, free intra-abdominal air, and appendiceal wall destruction. After internal validation by applying a shrinkage factor of 0.906, the final model demonstrated a C statistic of 0.87 (95% CI, 0.84-0.89) (eFigure 2 in Supplement 1). The SAS 2.0 version without CRP resulted in a C statistic of 0.83 (95% CI, 0.80-0.86) (eAppendix, eTable 3, and eFigure 4 in Supplement 1).

Validation of the SAS 2.0

The SAS 2.0 model was externally validated in a cohort of 565 patients, 184 of whom (32.6%) had complicated appendicitis. A C statistic of 0.86 (95% CI, 0.82-0.89) was found (eFigure 2 in Supplement 1). The calibration curve in the external validation cohort followed closely along the diagonal line, indicating that estimated risks corresponded well to observed proportions. Validation of the SAS 2.0 version without CRP as a predictor resulted in a C statistic of 0.82 (95% CI, 0.78-0.86).

When the SAS 2.0 was used in a subgroup of only patients diagnosed by US, a C statistic of 0.82 (95% CI, 0.76-0.87) was seen, vs 0.85 (95% CI, 0.79-0.90) in the subgroup of only patients diagnosed by CT. This suggests that the accuracy of SAS 2.0 is comparable in patients who have been diagnosed with US and CT and can be used as a single score independent of the imaging modality used.

SAS 2.0 Web Application for Individual Assessment of Probability of Complicated Appendicitis

The validated SAS 2.0 was implemented in a web application that can be found at http://www.sasappendicitis.com, where a patient's probability of having complicated appendicitis can be calculated. The sensitivity and NPV of the SAS 2.0 can be derived from eFigure 3 in Supplement 1.

Figure 2 illustrates the clinical application of the SAS 2.0 using 3 hypothetical patient scenarios. The first patient has a low probability of complicated appendicitis, the second a moderate probability, and the third a high probability. SAS 2.0 calculates the probability of complicated appendicitis individually for each patient, offering a tailored and accurate assessment. Hypothetical patients 2 and 3, despite having markedly different probabilities of complicated appendicitis, would have been lumped together if a cutoff value had been used.

Table 3. Univariable and Multivariable Logistic Regression Analysis of the Development Cohort (n = 795)

Variable[a]	Odds ratio (95% CI)	
	Univariable model	Multivariable model[b]
Male sex	1.62 (1.21-2.18)	1.65 (1.09-2.33)
Age, y[c]	2.95 (2.27-3.83)	2.00 (1.44-2.77)
Body temperature, °C[c,d]	1.73 (1.42-2.11)	1.19 (0.92-1.53)
NPRS score[c]	1.40 (1.11-1.77)	1.57 (1.15-2.15)
Vomiting	1.22 (0.89-1.66)	NA
Duration of symptoms, d[c]	1.96 (1.50-2.57)	NA
White blood cell count, µL[c]	1.52 (1.25-1.85)	1.28 (0.98-1.66)
C-reactive protein, mg/dL[c]	3.41 (2.74-4.24)	2.33 (1.83-2.96)
As found by imaging (computed tomography or ultrasonography)		
Appendiceal diameter, mm[c]	2.30 (1.84-2.86)	1.46 (1.11-1.92)
Free intra-abdominal fluid	2.49 (1.85-3.35)	1.52 (1.03-2.23)
Intra-abdominal abscess	18.0 (5.41-59.9)	4.81 (1.17-19.8)
Appendicolith	2.86 (2.10-3.89)	2.10 (1.41-3.13)
Fat infiltration	7.41 (2.72-20.2)	4.70 (1.24-17.8)
Free intra-abdominal air	20.4 (7.22-57.8)	3.40 (1.02-11.4)
Appendiceal wall destruction	3.86 (2.39-6.23)	2.01 (1.04-3.89)

Abbreviations: NA, not applicable; NPRS, numeric pain rating scale (score from 0-10).

[a] Missing values were imputed using 25 datasets. Missing percentages were 16% for NPRS, 6% for vomiting, 10% for appendiceal diameter, and 7% for fat infiltration; all other variables were complete.

[b] Odds ratios from the multivariable model are described after internal validation by applying a shrinkage factor of 0.906 based on a bootstrap resampling with 500 replicates.

[c] Odds ratios for continuous variables represent IQR odds ratios (IQR for age, 29-57 years; temperature, 36.7-37.7 °C; NPRS score, 4-7; duration of symptoms, 1-2 days; white blood cell count, 10 300-16 100 µL (to convert to $\times 10^9$/L, multiply by 0.001); C-reactive protein, 2.1-10.1 mg/dL (to convert to mg/L, multiply by 10); appendiceal diameter, 9-13 mm.

[d] The variable body temperature was modeled using restricted cubic splines.

Discussion

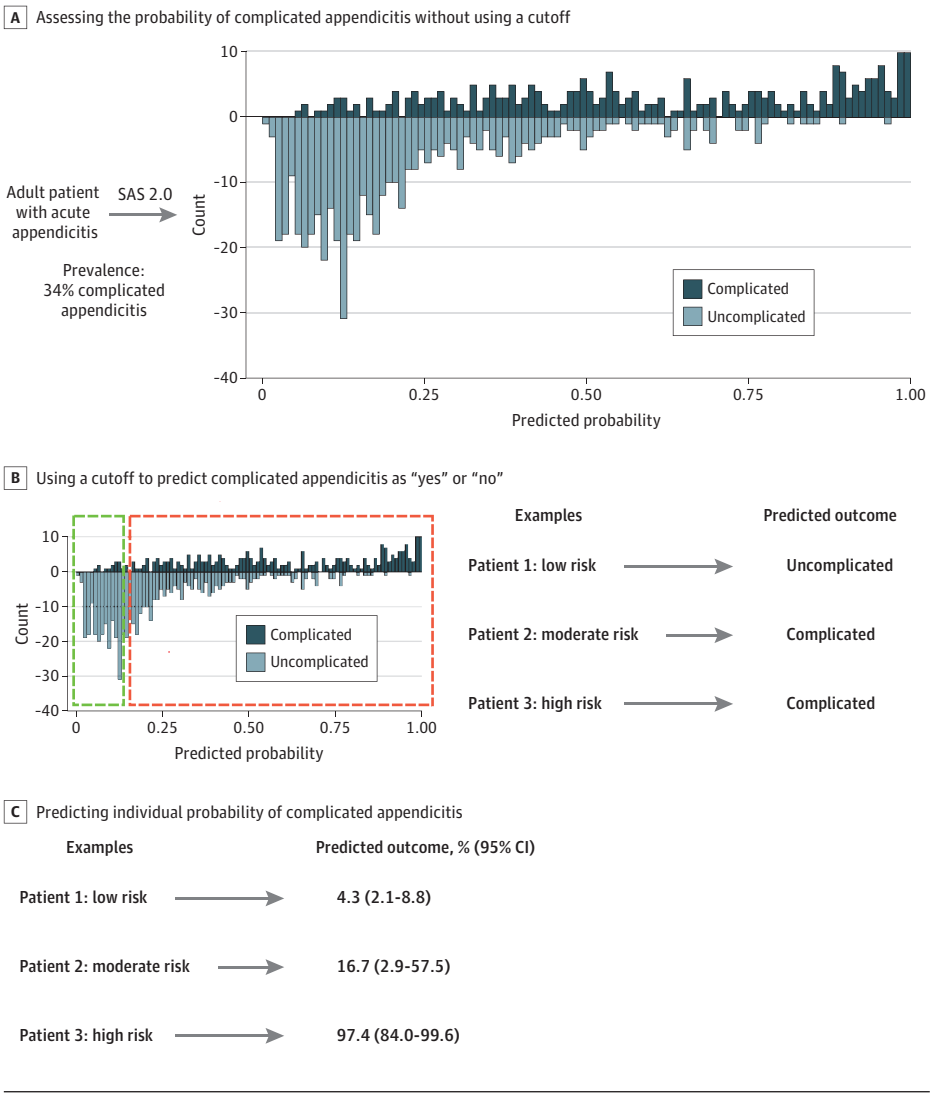
This prospective study found that the SAS[14] was not accurate enough to rule out complicated appendicitis. Therefore, the SAS 2.0 was developed and validated. In patients with imaging-confirmed acute appendicitis, the SAS 2.0 was able to accurately calculate an individual patient's probability of complicated appendicitis. An online tool was designed to enable implementation of the SAS 2.0 in clinical practice.

The SAS 2.0 could be used to select patients for nonoperative treatment, as only patients with uncomplicated appendicitis can be expected to benefit from nonoperative therapy. Some chance of complicated appendicitis could be accepted, as within trials demonstrating that antibiotics are a safe treatment option for acute appendicitis, 16.9% of patients were found to have complicated appendicitis.[7] Use of the SAS 2.0 as a selection criterion could lead to even better results in the nonoperative treatment of appendicitis. However, it should

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Copenhagen University Library, Gabriel Wentz on 03/28/2024

Figure 2. The Scoring System of Appendicitis Severity (SAS) 2.0



The SAS 2.0 assesses the probability of having complicated appendicitis for patients with acute appendicitis without using a cutoff (A). Use of a cutoff would result in a binary outcome wherein the outcome is highly influenced by the chosen cutoff value and many different patients are grouped together under the same denominator (B). SAS 2.0 computes a probability along with a confidence interval, enabling physicians to properly inform their patients and together make a decision regarding further treatment options (C). This way, the physician and patient can determine together what risk of complicated appendicitis is acceptable.

be noted that the SAS 2.0 only provides information on the likelihood of complicated appendicitis and does not guarantee the success of nonoperative treatment.

When considering nonoperative treatment of acute appendicitis, surgeons and patients need to carefully evaluate the potential risks of missing a diagnosis of complicated appendicitis or failure of antibiotic treatment. In doing so, some physicians and patients are willing to take greater risks than others to avoid initial surgery.[26-28] Hence, in contrast to our original view when developing the SAS, a predetermined cutoff value to identify uncomplicated or complicated appendicitis may not be applicable to most patients. The SAS 2.0 provides a patient-specific probability of complicated appendicitis among patients with acute appendicitis, allowing for personalized treatment consideration. To our knowledge, the only previous model that provides an individual risk assessment for complicated appendicitis such as SAS 2.0 was developed by Kim et al.[12] However, users of their model must enter the prevalence of complicated appendicitis within their own hospital as well as a target sensitivity, making it less useful to the average physician. Moreover, automatically integrating a cutoff point in the model from Kim et al results in a recommended treatment of whether or not to perform an appendectomy, which does not account for patient preference. We believe that clinicians should make well-informed treatment choices together with their patients. For this reason, we present only the risk of having complicated appendicitis with a 95% CI and do not provide any treatment recommendation with the SAS 2.0.

We found that the percentage of patients with complicated appendicitis within patients diagnosed with CT was significantly higher than in the US group (49.7% vs 25.1%, respectively). US with conditional CT is the standard approach for the diagnosis of acute appendicitis in the Netherlands,[29] but in critically ill patients with abdominal pain, CT is often performed immediately to rule out other abdominal causes, which may have resulted in more cases of complicated appendicitis in the CT group. To develop a score that could be used in all patients worldwide, appropriate radiological parameters were

© 2024 American Medical Association. All rights reserved.
Downloaded from jamanetwork.com by Copenhagen University Library, Gabriel Wentz on 03/28/2024

included regardless of their being detected by CT or US. Although this generalization improves clinical applicability, it may also lead to some bias, as CT is more accurate for finding free intra-abdominal air or fluid than US.

The reference standard we used is common and largely consistent with that of Atema et al[14] but deviates from our study protocol.[15] The protocol diagnosed cases of appendicitis with microscopic alterations consistent with complicated appendicitis as being complicated appendicitis, even when intraoperative findings rated the appendicitis as noncomplicated. This would have resulted in a rate of complicated appendicitis of 45.1%. This overdiagnosis would not be consistent with previous publications. We favored a practice-based reference standard when discrepancy between histopathological and intraoperative findings occurred and placed greater reliance on the surgeons' judgment. This adjustment brought present complicated appendicitis proportion more closely to previous studies (34.6% vs 45.1%, respectively). In the stricter reference standard that was initiated in the study protocol, most extracomplicated appendicitis diagnoses relied on histological gangrene (97.0%) without intraoperative signs of necrosis, and all but 1 were microscopic gangrene. Another deviation from the study protocol was the omission of a predetermined cutoff point. This decision was motivated by the growing realization that a universal cutoff point may not account for patients' and physicians' preferences[30] and would ignore the importance of conducting shared-decision making with patients. Accuracy of the SAS 2.0 cannot be determined in terms of sensitivity, NPV, specificity, or PPV, as we decided against the use of a fixed cutoff value.

Limitations

This study has limitations. First, a large proportion of patients were diagnosed by US, so the model may perform differently in a population where only CT is used. Furthermore, a considerable proportion of potential participants were not included. This could be attributed to the study's implementation across 11 hospitals, posing challenges for rigorous oversight, and potentially leading to a failure to approach certain patients for participation.

Conclusion

This was a prospective multicenter study. The study protocol was published before starting the final analysis, and the sample size was met. Only real-life data were used, leading to high applicability of the model. The analysis primarily focuses on accurately ruling out complicated appendicitis, as this is crucial when nonoperative treatment is considered. Correctly ruling in complicated appendicitis had lower priority in the setup, as patients with true uncomplicated appendicitis who are wrongly labeled as having complicated appendicitis will undergo appendectomy as treatment and will not experience undertreatment.[5] However, the relevance of correctly identifying complicated appendicitis should not be overlooked. Patients with true complicated appendicitis require urgent appendectomy to reduce the risk of postoperative complications.[31]

External validation did not find the original SAS to be sufficiently accurate in excluding complicated appendicitis. Therefore, the SAS 2.0 was developed to accurately calculate an individual patient's probability of having complicated appendicitis based on clinical, biochemical, and imaging findings among adults with acute appendicitis. The online SAS 2.0 tool allows for easy application of the prediction model in practice and can provide a reasonable basis for when to consider and discuss nonoperative treatment of acute appendicitis with patients.

ARTICLE INFORMATION

Accepted for Publication: December 26, 2023.

Published Online: March 27, 2024. doi:10.1001/jamasurg.2024.0235

Author Affiliations: Department of Surgery, Amsterdam University Medical Center, Amsterdam, the Netherlands (Scheijmans, Bom, Ghori, Boermeester); Department of Surgery, Tergooi Medical Center, Hilversum, the Netherlands (van Geloven, van de Wouw); Department of Medical Imaging, Radboud University Medical Center, Nijmegen, the Netherlands (Hannink); Department of Surgery, Maasstad Hospital, Rotterdam, the Netherlands (van Rossem); Department of Radiology, Tergooi Medical Center, Hilversum, the Netherlands (Huisman); Department of Surgery, Spaarne Gasthuis, Haarlem and Hoofddorp, the Netherlands (van Hemert, Franken, Oosterling); Department of Surgery, Radboud University Medical Center, Nijmegen, the Netherlands (Rosman); Department of Pathology, Amsterdam University Medical Center, Amsterdam, the Netherlands (Koens); Department of Radiology and Nuclear Medicine, Amsterdam University Medical Center, Amsterdam, the Netherlands (Stoker); Amsterdam Gastroenterology Endocrinology and Metabolism, Amsterdam, the Netherlands (Stoker, Boermeester); Department of Epidemiology and Data Science, Amsterdam University Medical Center, Amsterdam, the Netherlands (Dijkgraaf); Methodology, Amsterdam Public Health, Amsterdam, the Netherlands (Dijkgraaf).

Author Contributions: Drs Scheijmans and Boermeester had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Scheijmans, Bom, Ghori, van Geloven, Hannink, van Rossem, Huisman, Oosterling, Stoker, Dijkgraaf, Boermeester.
Acquisition, analysis, or interpretation of data: Scheijmans, Bom, Ghori, van Geloven, Hannink, van Rossem, van de Wouw, van Hemert, Franken, Oosterling, Rosman, Koens, Boermeester.
Drafting of the manuscript: Scheijmans, Bom, Ghori, Hannink.
Critical review of the manuscript for important intellectual content: Scheijmans, Bom, van Geloven, Hannink, van Rossem, van de Wouw, Huisman, van Hemert, Franken, Oosterling, Rosman, Koens, Stoker, Dijkgraaf, Boermeester.
Statistical analysis: Scheijmans, Bom, Ghori, Hannink, Dijkgraaf.
Obtained funding: Bom, Stoker, Dijkgraaf, Boermeester.
Administrative, technical, or material support: Scheijmans, Bom, Ghori, van Geloven, van de Wouw, van Hemert, Koens, Boermeester.
Supervision: Bom, van Geloven, Hannink, van Rossem, Oosterling, Rosman, Koens, Boermeester.

Conflict of Interest Disclosures: Drs Scheijmans and Stoker reported grants from Maag Lever Darm Stichting during the conduct of the study. Dr Boermeester reported institutional fees for serving as a speaker and/or instructor from 3M, Johnson & Johnson, Gore, Smith & Nephew, Angiodynamics, Telabio, Medtronic, and Molnlycke outside the submitted work. No other disclosures were reported.

Funding/Support: This study was supported by Maag Lever Darm Stichting grant D18-05.

Role of the Funder/Sponsor: The funder was not involved in the collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: The members of the SAS Collaborative Group appear in Supplement 2.

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Copenhagen University Library, Gabriel Wentz on 03/28/2024

Data Sharing Statement: See Supplement 3.

REFERENCES

1. Salminen P, Paajanen H, Rautio T, et al. Antibiotic therapy vs appendectomy for treatment of uncomplicated acute appendicitis: the APPAC randomized clinical trial. *JAMA*. 2015;313(23):2340-2348. doi:10.1001/jama.2015.6154

2. Vons C, Barry C, Maitre S, et al. Amoxicillin plus clavulanic acid versus appendicectomy for treatment of acute uncomplicated appendicitis: an open-label, non-inferiority, randomised controlled trial. *Lancet*. 2011;377(9777):1573-1579. doi:10.1016/S0140-6736(11)60410-8

3. Flum DR, Davidson GH, Monsell SE, et al; CODA Collaborative. A randomized trial comparing antibiotics with appendectomy for appendicitis. *N Engl J Med*. 2020;383(20):1907-1919. doi:10.1056/NEJMoa2014320

4. O'Leary DP, Walsh SM, Bolger J, et al. A randomised clinical trial evaluating the efficacy and quality of life of antibiotic only treatment of acute uncomplicated appendicitis: results of the COMMA trial. *Ann Surg*. 2021;274(2):240-247. doi:10.1097/SLA.0000000000004785

5. Bom WJ, Scheijmans JCG, Salminen P, Boermeester MA. Diagnosis of uncomplicated and complicated appendicitis in adults. *Scand J Surg*. 2021;110(2):170-179. doi:10.1177/14574969211008330

6. Skjold-Ødegaard B, Søreide K. The diagnostic differentiation challenge in acute appendicitis: how to distinguish between uncomplicated and complicated appendicitis in adults. *Diagnostics (Basel)*. 2022;12(7):1724. doi:10.3390/diagnostics12071724

7. Rollins KE, Varadhan KK, Neal KR, Lobo DN. Antibiotics versus appendicectomy for the treatment of uncomplicated acute appendicitis: an updated meta-analysis of randomised controlled trials. *World J Surg*. 2016;40(10):2305-2318. doi:10.1007/s00268-016-3561-7

8. Imaoka Y, Itamoto T, Takakura Y, Suzuki T, Ikeda S, Urushihara T. Validity of predictive factors of acute complicated appendicitis. *World J Emerg Surg*. 2016;11:48. doi:10.1186/s13017-016-0107-0

9. Khan MS, Siddiqui MTH, Shahzad N, Haider A, Chaudhry MBH, Alvi R. Factors associated with complicated appendicitis: view from a low-middle income country. *Cureus*. 2019;11(5):e4765. doi:10.7759/cureus.4765

10. Kim TH, Cho BS, Jung JH, Lee MS, Jang JH, Kim CN. Predictive factors to distinguish between patients with noncomplicated appendicitis and those with complicated appendicitis. *Ann Coloproctol*. 2015;31(5):192-197. doi:10.3393/ac.2015.31.5.192

11. Lin HA, Tsai HW, Chao CC, Lin SF. Periappendiceal fat-stranding models for discriminating between complicated and uncomplicated acute appendicitis: a diagnostic and validation study. *World J Emerg Surg*. 2021;16(1):52. doi:10.1186/s13017-021-00398-5

12. Kim HY, Park JH, Lee SS, Jeon JJ, Yoon CJ, Lee KH. Differentiation between complicated and uncomplicated appendicitis: diagnostic model development and validation study. *Abdom Radiol (NY)*. 2021;46(3):948-959. doi:10.1007/s00261-020-02737-7

13. Avanesov M, Wiese NJ, Karul M, et al. Diagnostic prediction of complicated appendicitis by combined clinical and radiological appendicitis severity index (APSI). *Eur Radiol*. 2018;28(9):3601-3610. doi:10.1007/s00330-018-5339-9

14. Atema JJ, van Rossem CC, Leeuwenburgh MM, Stoker J, Boermeester MA. Scoring system to distinguish uncomplicated from complicated acute appendicitis. *Br J Surg*. 2015;102(8):979-990. doi:10.1002/bjs.9835

15. Bom WJ, Scheijmans JCG, Ubels S, et al. Optimising diagnostics to discriminate complicated from uncomplicated appendicitis: a prospective cohort study protocol. *BMJ Open*. 2022;12(4):e054304. doi:10.1136/bmjopen-2021-054304

16. Geerdink TH, Augustinus S, Atema JJ, Jensch S, Vrouenraets BC, de Castro SMM. Validation of a scoring system to distinguish uncomplicated from complicated appendicitis. *J Surg Res*. 2021;258:231-238. doi:10.1016/j.jss.2020.08.050

17. Lastunen K, Leppäniemi A, Mentula P. Perforation rate after a diagnosis of uncomplicated appendicitis on CT. *BJS Open*. 2021;5(1):zraa034. doi:10.1093/bjsopen/zraa034

18. Fujiwara K, Abe A, Masatsugu T, Hirano T, Hiraka K, Sada M. Usefulness of several factors and clinical scoring models in preoperative diagnosis of complicated appendicitis. *PLoS One*. 2021;16(7):e0255253. doi:10.1371/journal.pone.0255253

19. Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527. doi:10.1136/bmj.h5527

20. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Ann Intern Med*. 2015;162(10):735-736. doi:10.7326/L15-5093-2

21. Bhangu A, Søreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. *Lancet*. 2015;386(10000):1278-1287. doi:10.1016/S0140-6736(15)00275-5

22. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441. doi:10.1136/bmj.m441

23. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230. doi:10.1186/s12916-019-1466-7

24. Laméris W, van Randen A, Dijkgraaf MG, Bossuyt PM, Stoker J, Boermeester MA. Optimization of diagnostic imaging use in patients with acute abdominal pain (OPTIMA): design and rationale. *BMC Emerg Med*. 2007;7:9. doi:10.1186/1471-227X-7-9

25. Leeuwenburgh MM, Wiarda BM, Wiezer MJ, et al; OPTIMAP Study Group. Comparison of imaging strategies with conditional contrast-enhanced CT and unenhanced MR imaging in patients suspected of having appendicitis: a multicenter diagnostic performance study. *Radiology*. 2013;268(1):135-143. doi:10.1148/radiol.13121753

26. Bom WJ, Scheijmans JCG, Gans SL, Van Geloven AAW, Boermeester MA. Population preference for treatment of uncomplicated appendicitis. *BJS Open*. 2021;5(4):zrab058. doi:10.1093/bjsopen/zrab058

27. Hanson AL, Crosby RD, Basson MD. Patient preferences for surgery or antibiotics for the treatment of acute appendicitis. *JAMA Surg*. 2018;153(5):471-478. doi:10.1001/jamasurg.2017.5310

28. Reinisch A, Reichert M, Hecker A, Padberg W, Ulrich F, Liese J. Nonoperative antibiotic treatment of appendicitis in adults: a survey among clinically active surgeons. *Visc Med*. 2020;36(6):494-500. doi:10.1159/000506058

29. Dutch Society for Surgery. Richtlijn acute appendicitis. Acute appendicitis guideline. Accessed February 21, 2024. https://richtlijnendatabase.nl/richtlijn/acute_appendicitis/startpagina_-_acute_appendicitis.html

30. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:i6. doi:10.1136/bmj.i6

31. Bolmers MDM, de Jonge J, Bom WJ, van Rossem CC, van Geloven AAW, Bemelman WA; Snapshot Appendicitis Collaborative Study group. In-hospital delay of appendectomy in acute, complicated appendicitis. *J Gastrointest Surg*. 2022;26(5):1063-1069. doi:10.1007/s11605-021-05220-w

© 2024 American Medical Association. All rights reserved.

Downloaded from jamanetwork.com by Copenhagen University Library, Gabriel Wentz on 03/28/2024